U.S. FARMERS' DECISION TO CHOOSE DIRECT SALES CHANNEL: A FRACTIONAL SEEMINGLY UNRELATED SEMIPARAMETRIC MODEL

Krishna P. Paudel, Timothy A. Park, and Mahesh Pandit[1]

Corresponding Author:

Krishna P. Paudel

Professor

Department of Agricultural Economics and Agribusiness

Louisiana State University (LSU) and LSU Agricultural Center

Baton Rouge, LA 70803

Phone: (225) 578-7363

Fax: (225) 578-2716

Email: kpaudel@agcenter.lsu.edu

[1] Paudel and Pandit are professor and former graduate student, respectively at Louisiana State University and LSU AgCenter, Baton Rouge, LA 70803. Park is a branch chief at the Farm and Rural Business Branch, Economic Research Service, USDA. The views expressed in the paper are of the authors and should not be attributed to the Economic Research Service or the U.S. Department of Agriculture. The authors would like to thank excellent research help from Seydina Sene.

U.S. FARMERS' DECISION TO CHOOSE DIRECT SALES CHANNEL: A FRACTIONAL SEEMINGLY UNRELATED SEMIPARAMETRIC MODEL

ABSTRACT

This paper examines the dynamic structure of retail food system in the United States and choice of direct marketing (DM) sales channel using cross-sectional data available from the United State Department of Agriculture Agricultural Resource Management Survey and North American Industry Classification System. We first clustered the data into several homogenous groups and conducted a fractional seemingly unrelated semiparametric model to identify variables affecting volumes of sales to different marketing channels. We also use a multinomial logistic regression model to identify factors affecting farmers' choice of direct marketing of crops and livestock products using farmers markets and on-farm stores (Channel 1), retails outlets (Channel 2), and both channels 1&2 (Channel 3). Our base category is when farmers choose not to engage in direct marketing (no direct sales). We find that farmers located in the Northeast regions are more likely to engage in direct marketing than farmers who are located in the Midwest, Pacific, and Mountain states. We also find that small farmers are more likely to be involved in direct marketing. The average probability of using no direct sales in the United States is 94% higher than direct marketing options combined ceteris paribus.

Keywords: channels choice, cluster analysis, direct marketing, fractional regression, grain, livestock, multinomial logit, seemingly unrelated fractional semiparametric model

JEL Classification: C25, Q12, Q13, Q16

U.S. FARMERS' DECISION TO CHOOSE DIRECT SALES CHANNEL: A FRACTIONAL SEEMINGLY UNRELATED SEMIPARAMETRIC MODEL

I. INTRODUCTION

The objective of this study is to understand the structural changes in retail food systems in which farmers sell crops and livestock products directly to consumers by choosing various marketing channel options in the United States. Farmers' choice of direct marketing channels is modeled first clustering the data using a mixture of expert model and then analysing it using a fractional seemingly unrelated semiparametric model. We also supplement the analyses using a multinomial logit regression model.

In the United States, the number of farmers engaged in direct marketing grew from about 340 in 1970 to over 3000 in 2001 and dramatically increased after the passing of the U.S. Public Law 94-463 (PL 94-463), the farmer-to-consumer Direct Marketing Act of 1976 (Brown 2002). Moreover, the number of farmers adopting direct-marketing strategies (DMS) has increased by 17 percent from 2002 to 2007 (Detre et al. 2011). Direct marketing has been enjoying per capita spending of $4.17, and an average growth in sales of 1.63% per year since 2007 (Boys and Blank 2016). Uematsu and Mishra (2011) indicated that the economic incentives available for both producers and consumers have contributed to the recent trend in increased use of direct marketing strategies by U.S farmers.

Several authors (Kohls and Uhl 1998, Buhr 2004) have looked at choice of marketing channels like pick-your-own, catalogue, and Internet sales operations,

consumer cooperatives agriculture (CSA), and locally branded meats by focusing on the consumers side (see, Brown et al. 2006, Monson et al. 2008). Uematsu and Mishra revealed that there are relatively fewer studies on the production side of direct marketing strategies. Moreover, Park (2015) posits that producers who schedule to sell more by adopting local outlets should expect to lose market share but this decline in sales is not proportional to the marginal increase in sales (Park et al. 2014, Low and Vogel 2011, Park, 2009, Darby et al. 2008, Schneider and Francis 2005). These empirical evidence motivate this paper to investigate factors determining the choice of direct marketing channels in the retail food system in the United States.

We analyze existing retail structure which has profound implication in estimating earning performance by considering farmer financial performances as a proxy for unobserved covariates. The model adopted here is analogous to a logistic regression model, except that the probability distribution of the response variable is multinomial instead of binomial. Moreover, the model is simpler and in line with the model environment adopted by Gourieroux and Monfort (1996).

Angelo et al. (2016) posit that the vertical nature of food supply chains makes it difficult for small and mid-sized farmers to access markets that could provide a significant return on investment. This paper addresses the concern that Angelo et al. (2016) raised regarding the lack of specific quantitative values for critical financial data.

We find that farmers located in the Northeast are more likely to engage in direct marketing than farmers who are located in the Midwest, Pacific, and Mountain

regions in the United States. This finding is vital and highlights significant and important regional differences in local foods markets that policymakers in the United States should take into consideration. This finding is consistent with Boys and Blank (2016) who indicate that consumers in U.S. states of Maine, New Hampshire, and Vermont purchase notably more from farmers than do consumers elsewhere in the country. They further found that the Southeast part of the United States lags particularly behind the national average when it comes to direct purchase of farm products.

We suggest that while government and non-government organization programs are emerging to support the marketing of local foods in the United States, policymakers should pay particular attention to regional differences. This will help policymakers to better channel efforts and find the promotion of local food to support agriculture producers in the United States.

The rest of the paper is organized as follow. Section II reviewed the literature on direct marketing. Section III describes the data collection procedure and calculation methods. Section IV provides the model structure linking the observables parameters and unobservable others agglomeration effects under certain parametric forms and distributional assumptions, and normalization process. Section V discusses the results. Section VI concludes.

II. LITERATURE REVIEW

We review factors driving or motivating consumers to engage in direct marketing. We

also provide justifications of market outlet choices influenced by the existing county spatial relationships that characterize the producers' behavior as well as their proximity to viable direct markets.

The existing literature on direct marketing channel outlet has mainly focused on the factors driving or motivating consumers to purchase directly from farmers. For instance, Wolf (1997) argues that the majority of consumers in San Luis Obispo County, California preferred produce at farmers markets rather than supermarkets because of tastes (freshness), quality, and price. Brown et al. (2006) use OLS model and test the relationship between county-level direct marketing sales and socio-economic, output and location characteristic focusing in West Virginia. They find that higher median housing value and population density among the young, location and distance to Washington, D.C. increases direct marketing (DM) adoption. First, these findings highlight the influence that location infrastructure, distance to metropolitan areas like Washington, D.C has on direct marketing outlets. Second, this paper also adds that the spillover effect of political awareness via regulations across industry can further empower consumers to make more informed choices about food consumption. Gallons et al. (1997) posit that consumers in Delaware choose direct market outlets because of its local and diverse option of locally grown products and the fact that consumers themselves directly boost producer's welfare. However, the results are only relevant for states with a large proportion of small part-time farmers close to agglomeration in metropolitan areas. Nevertheless, we find that there is a gap in understanding the determinants of the decision to engage in direct marketing

choice. Monson et al. (2008) document that there is a dearth of literature on the determinants of the decision to direct market. However, they find that factors such as farm size, high-value crops, organic product, farmer experience, and distributional infrastructure affect a producer's choice of the direct markets outlet channel. One of the hypotheses put forward in the literature is that as the farm size increases, farmers will no longer choose direct marketing channels outlets. The second hypothesis is that as the share of high-value products in the farm portfolio increases, farmers are less willing to choose direct marketing channels outlets. They, however, do not find any evidence on their relationship with farmers' earning performance especially nationwide and it is only relevant for the share of high-value products in the farm portfolio and among more educated farmers. Moreover, as direct marketers take on the role of large store food retailers, county control influences on agglomeration and choice of direct marketing extending to farmers earning are paramount for policymakers to understand and formulate policies in the United States.

Detre et al. (2011) use a double-hurdle model and the 2002 Agricultural Resource Management Survey to evaluate the adoption of direct marketing strategies (DMS) and their impacts on farm bottom line in the United States. They find that the spatial parameters of the farm constraint by the changing structure of consumer's preferences in organic food production positively affect the adoption and structure of direct marketing strategies (DMS) without mentioning selectivity bias. Park (2015) uses the unconditional quantile regression estimator to investigate the impact of

participation in direct marketing on the entire distribution of farm sales and found that the heterogeneous effects take place across the distribution of farm sales. Uematsu and Mishra (2011) estimate a zero-inflated negative binomial model to identify factors influencing the total number of direct marketing strategies adopted by farmers and find that the degree of intensity of the adoption has no significant impact on gross cash farm income without addressing the selectivity bias. Our observation is that most of the studies have regional or state focus and those only control for the impact of adopting different direct marketing outlets. Uematsu and Mishra (2011) fill the gap using the national survey controlling for farm sizes but found no effects. Instead, they conclude that adoption of individual direct marketing strategies showed some significant effects on gross cash farm income. Our contribution is thus to include farmers financial performance and its influence on the choice of direct marketing.

Another important feature of direct marketing is drawing from the literature of urban economics linking agglomeration as a place marketing site and the structural and behavioral conditions present within alternative market channel outlets. In this literature, retail locations have been characterized by common preferences of clustering among retailers to create an opportunity for place marketing activities (Teller and Reutterer, 2008). Thus, consumers, producers, and markets do affect the structure of direct marketing location. Jarosz (2008) enumerates a list of county infrastructural parameters influencing market channel outlet such as (1) distances between producer and consumer, (2) farm size and scale, (3) whether

organic or holistic, (4) agribusiness-orientation, (5) alliance cooperatives, and (6) and commitment to the social (social justice / equity concerns), economics' activism and environmental considerations of food production.

These infrastructural parameters illustrate what Guthman et al. (2006) explain as region or country's agricultural geography and history which play important roles in production and farms location. For instance, Sage (2012) investigates the geographical exploration and economic factors of farmers markets compared to those in rural communities and found that in Washington State the western district contains 65 percent of all the farmer's markets in and the Western District has the lowest per capita number of markets of any of the districts. This indicates a potential pull factor in engaging in direct marketing in term of attractiveness (Hay and Smith, 1980) and a lack of a market and consumer base could make farmer participate less in direct marketing regardless of sales opportunities and performance. In this literature, distance or the radius of trade are the most influential factor that affects the final decision for market outlets strategies which leads to their agglomeration in a specific county. The argument is that such agglomeration enables the use of common infrastructure and environment and provides access to consumers regardless of the purchasing power, let alone retailers' total sales (Teller and Reutterer, 2008). For instance, Govindasamy et al. (1998) find that farmers in New Jersey travel 1-70 miles to access direct market. Metcalf (1999) reports that Dane Country, Wisconsin farmers' market drew farmers and customers from a radius of 240 miles.

Another important parameter to control for county is age which is considered as one of the most important parameters in the configuration of the customer base network of a farmers choice of location and direct marketing outlets: The hypothesis is that farmers draw naturally from the neighbourhoods nearby where they are located and most of the customers in farmers' market, for instance, are Caucasian (if race is reported) (Capstick, 1982). Among factors that influence the decision to adopt direct marketing strategies regardless of earning performance is whether farmers are an organic producer or not. The argument is that farmers who adopt organic markets are required to have a third-party certification of compliance with the USDA's National Organic Standard (NOS) if sales are greater or equal to $5,000 (Monson et al. 2008). Thus, if farmers are organic producers and sales is less than $5000 farmers have no incentive to seek the USDA's National Organic Standard compliance, and therefore regardless of earning potential they will adopt direct marketing outlets. Gender is also another factor to consider because female farmers are more inclined to adopt direct greater marketing channel outlets. The argument is the ratio of women to men in high-value markets including direct marketing channel is higher than in commodity markets (Monson et al. 2008). Education level is also another influencing factor to consider given that as education increases adoption of direct marketing outlets increases. The argument is as follow: farmers who have higher earning potential in off-farm activities is more inclined to invest in high-value product and thus naturally adopt direct marketing outlets given that they usually buy rural land with the purpose of farming as a hobby or retirement activity (Monson et al. 2008).

Therefore regardless of earnings potential from direct marketing, these farmers depend on their location in a county and off-farm activity are not motivated by earning performance when choosing direct in marketing outlets.

III. DATA

This paper uses data obtained from the nationwide 2011 Agricultural Resource Management Survey (ARMS) collected by the USDA Economic Research Service (ERS) and the National Agricultural Statistics Service (NASS) in the United States. On the demand side, we got the retail structure of each county such as local grocery stores, restaurants, and others retailers where the respondent is located. We combined USDA ARMS data with the NAICS data. We use four sectors in the global retail industries, *viz.*, (1) Grocery and related product merchant wholesalers (NAICS 4244); (2) Farm product raw material merchant wholesalers (NAICS 4245); (3) Groceries stores (NAICS 4451); (4) Restaurants and others eating places (NAICS 7225). The United States Census Bureau provides information about a total number of establishments, employee class size, total industry payroll by NAICS and annual payroll ($1,000). To calculate the total size of employment, annual payroll, and the number of establishment in all industries used in this paper, we use a midpoint for employment in each category. We merge total employment across industries and by size class to obtain final numbers for all counties in 2011.

The ARMS data provided information relative to the link between total agricultural production, farms characteristics, and financial profiles. To determine market channel outlets, we utilize information provided by farmers regarding

different channels used for direct marketing outlets (DM) such as roadside stand or, on-farm store, farmers market, community supported agriculture (CSA), regional distributors, state branding programs, direct sales to local grocery stores, restaurants, super centers, convenience stores. The descriptive statistics are presented in Table 2.

Regional dummy variables are created and included as explanatory variables in the model. Mountain is when farm is located in one of the Mountain states (1=yes, 0 otherwise); Northeast is when farm is located in one of the Northeast states (1=yes, 0 otherwise), Midwest is when farm is located in one of the Midwest states (1=yes, 0 otherwise), and Pacific when farm is located in one of the Pacific states (1 = yes, 0 otherwise). The relative sales declines reported for participation in direct marketing ranged from 22 to 27 percent for the Pacific, Mountain, Midwest, and North Central regions with a smaller decline of about 9 percent for Southern producers. TOTEMP is total employment in all four industries within the respondent's county. Total pay is total annual payroll for all four industry, and TOTEST is the number of total establishments or retails store in the respondent's county. TACRES is the total farm land in acres. GFI is gross farm income, YOB is the year operator began his/her operation on the farm, WAGL is wage expenses for other labor, and OPMH is operator miles from home. Definitions of all variables included in the model are presented in Table 1.

IV. MODEL

Multiple approaches to group U.S. farmers are available and may be based on climate, census region, or kinds of crop/livestock produced. Many studies (Park 2015) have used regional dummies or census dummies to isolate regional effects on direct marketing or production behavior. We believe this approach should be replaced by identification of homogenous groups based on characteristics and identified using statistical methods. If we group U.S. farmers with similar characteristics, and do analysis in each cluster, the results obtained from the analyses should be more valid. Given the available clustering procedures, results suggest that a model based cluster analysis performed better in specific applications (Gormley and Murphy, 2010).

Existing literature on market channel choice analysis has used a parametric method without due consideration of fraction of sales volumes to each channel. This clearly is a flawed analysis, as fractions of outputs going to each market channel (no direct channel and various options of direct channel) for a farm should sum to 100% and those sales volumes are related to each other. The estimated parameters from individual equations may not be consistent if channel choices are correlated. We address this concern using a multivariate fractional regression model. Further, if the distribution assumption used in the parametric regression is wrong, parameters will be inconsistent and inefficient. Hence, we use a nonparametric estimation procedure to capture the nonlinearity of variables in the model. We also note that previous studies assumed regional differences. As an alternative, we use a model

based cluster analysis to identify homogeneous groups of producers. We identify the impact of growers' business characteristics on shares of sales to these channels using a multivariate fractional regression model under each homogenous cluster group. We compare the parametric model to nonparametric using a model specification test developed by Hsiao et al. (2007).

Cluster Analysis

A heterogeneous population of U.S. farmers may be a collection of homogeneous subpopulations, but these subpopulations are unknown and can be characterized by appropriate clustering methods. We conduct a cluster analysis suggested by Gormley and Murphy (2010) in which clusters are formed based on explanatory variables. They developed a mixture of experts model (MoE) for the rank order data. Our data are not rank order, so the model is modified so that it can be used in a case when a dependent variable is fractional in nature. A mixture of experts model (Jacobs et al., 1991; Jordan and Jacobs, 1994) which combines the idea of mixture models (McLachlan and Peel, 2000) and generalized linear models (Gormley and Murphy, 2008) works well for fractional data.

As usual, the MoE model gives the relationship between dependent variable and explanatory variables, but this model assumes that the conditional distribution of the dependent variable given the explanatory variables is a finite mixture distribution (Gormley and Murphy, 2008). Let K be the total number of homogeneous subpopulations of a heterogeneous population also known as expert

networks. A gating network coefficient is the probability that an ith farmer belongs to subpopulation k which we designate as π_{ik} . The probability distribution for a farmer i being in subpopulation k is $p(V_i|\theta_k)$, where θ represents the parameters of the model for subpopulation k. Let X represents a matrix of associated explanatory variables for i^{th} farmer. Then the conditional probability of a farmer i's marketing channel choice V_i, given associated explanatory variables X is

$$P(V_i|X) = \sum_{k=1}^{K} \pi_{ik} P(V_i|\beta_k) \tag{1}$$

The gating network coefficients are weighting probabilities such that they are nonnegative and sum to one for each farmer. The probability of a farmer i's choice according to the expert networks in the mixture model are blended by the gating network coefficients to produce overall probability. Thus the probability of farmer i's preference is a convex combination of the output probabilities from the expert networks.

The gating network coefficients are assumed to be a function of the characteristics (explanatory variables) associated with U.S. farmers. These explanatory variables determine a particular cluster in which a nursery producer belongs. The gating network coefficients in the MoE can be estimated using a multinomial logistic function, since probability of U.S. farmers belonging to each of K expert networks can be viewed as success probabilities from a generalized linear model (Gormley and Murphy, 2008). Then a farmer i's gating network coefficients $\pi_i = (\pi_{i1}\pi_{i2}, \ldots, \pi_{iK})$ are modeled by a logistic function of their P explanatory variables $X = (x_{i1}, x_{i2}, \ldots, x_{iP})$. Then, the multinomial logistic model takes the following forms

(2) $$\log\left(\frac{\pi_{ik}}{\pi_{i1}}\right) = \theta_{k0} + \theta_{k1}x_{i1} + \theta_{k2}x_{i2} + \cdots + \theta_{kL}x_{iL}$$

where expert network 1 is used as the baseline expert network and θ are the gating network parameter estimates.

Fractional regression model

Within the context of market channel data, each expert network needs to be appropriately modelled. Assume a farmer chooses different market channels that make up the total portfolio of market shares. These different market channels are no direct marketing, direct marketing using farmers market or farm store, direct marketing using retail, and direct marketing using other institutions. A farmer therefore chooses to sell a fraction of output to a given market channel and the total sales volume going to the four market channels adds to 100% (or 1 given we are interpreting sales as a fraction here). Let $Y = (y_1, y_2, \ldots, y_m, \ldots, y_M)$ represent the fraction of output devoted to M different market channels. Since the values associated with these variables are fractions, they are limited to the closed interval [0,1]. An appropriate model should adjust the nature of fractional variables. A solution to deal with this type of variable is to use a nonlinear function satisfying $0 \leq g(.) \leq 1$, where $g(.)$ is nonlinear model proposed by Papke and Wooldridge (1996). Hence, the conditional mean of the dependent variable can be expressed as

(3) $$E(Y|X) = g(X\beta)$$

with X as a $(N \times P)$ matrix of independent variables and β as a vector of parameters. These independent variables are the same as used in determining the appropriate clusters. A fractional model is specified using logistic link with Bernoulli distribution. We estimate the β by maximizing Bernoulli log-likelihood function given by

$$LL(\beta) = \sum_{i=1}^{N} y_i \log[g_i(X\beta)] + (1 + y_i) \log[1 - g_i(X\beta)] \quad (4)$$

with N being the number of U.S. farmers. The estimated parameter is consistent and asymptotically normal provided that $E(Y|X)$ is correctly specified. Different approaches are discussed in previous literature for univariate cases (Papke and Wooldridge, 2008; Ramalho et al., 2011). These authors have proposed a fractional regression model on the basis of quasi-likelihood and logit conditional mean functions.

In our case, we estimate the model simultaneously using a multivariate specification as U.S. farmers choose four different channels and those channel choices are correlated. In a recent paper, Murteira et al. (2013) have proposed generalization of a univariate specification shown in equation (3) to a multivariate specification with multinomial logit link and multivariate Bernoulli distributions[2].

[2] An alternative to logit link function and Bernoulli distribution is to use a beta distribution in which density values lies between 0 and 1. However, this is less common compared to the quasi-likelihood maximum likelihood estimation. A paper by Ramalho et al. (2011) illustrated a different models and estimation procedure that can be used for multivariate fractional response variables with test procedure to check method validity.

Let $E(Y|X) = G(X;\beta) = [G_1(X,\beta_1), \dots, G_M(X,\beta_M)]'$ be the M vector of conditional mean function with its components $E(y_m|X), m = 1, \dots, M$, with $G_m = G_m(X,\beta_m)$. Here the conditional mean $0 < G_m < 1$ for all m and $\sum_{l=1}^{M} G_m = 1$.We use a multinomial logit[3] specification expressed as:

V. (5) $$G_m = \frac{\exp(X\beta_m)}{\sum_{l=1}^{M}\exp(X\beta_m)}, m = 1, \dots, M$$

Let y_m be the fraction of mth component market channels used by a nursery producer, then it follows the multivariate Bernoulli (MB) distribution (Murteira et al., 2013). So the individual contribution to the log-likelihood can be expressed as:

(6) $$\log L_i(\beta) = \sum_{m=1}^{M} y_{im} \log G_{im} = \sum_{m=1}^{M-1} \log \frac{G_{im}}{G_{iM}} + \log G_{iM}$$

Here, $G_{iM} = 1 - \sum_{m=1}^{M-1} G_{im}$. Then the quasi-maximum likelihood (QML) estimator is obtained by maximizing log-likelihood of all U.S. farmers (N) as given below:

(7) $$LL(\beta) = \sum_{i=1}^{N} \log L_i(\beta)$$

The estimated parameter $\hat{\beta}$ is consistent and asymptotically normal regardless of the true conditional distribution y, provided that G is correctly specified.

[3] We tested for the independence of irrelevant alternatives for the multinomial logit type of specification for fractional model using Murteira (2013)'s procedure and found that the assumption holds for our case. We followed method developed in that paper to calculate dissimilarity parameters (τ). Following that paper, the null hypothesis to test IIA assumption is H_0:$\tau_l = 0$. The τ values for all clusters are close to 0 (see below), and we fail to reject the null hypothesis in all conditions. This justifies our use of G function as a multinomial logit specification.

We estimated multivariate parametric and nonparametric fractional models for all four market channel choices within a cluster. To identify the suitability of a nonparametric specification, we used a test developed by Hsio et al. (2007).

Multinomial logit model

We also adopt a multinomial logit regression model. This approach allows us to model farmer's direct choice of marketing outlets among different options (categories) on several explanatory variables. In this paper, farmers face four marketing channel options: direct marketing of crops and livestock products using farmers markets and on-farm stores (Channel 1), retails outlets (Channel 2), and both channels 1&2 (Channel 3). Our base category is when farmers choose not to engage in direct marketing (no direct sales which is Channel 4).

In the multinomial logit setting, we assume that the log of odds of each choice or channel among the four categories in this paper follows a linear model characterized as follow:

$$\eta_{ij} = log\frac{\pi_{ij}}{\pi_{iJ}} = \alpha_J + X_i'\beta_j.$$

Here, α_J is a constant and β_j is a vector of regression coefficients for j=1, 2,. ..., J-1. However, it is very important to remember that the model matrix structure does not contains a column of ones. We only need J-1 to operationalize a variable with j choice categories regardless of the category we choose as our basis. In this paper we choose no direct sales as our base category against the other direct marketing channels.

Consequently, the multinomial logit model could be written in terms of the original probabilities π_{ij} instead of the log-odds specify as follow:

$$\Pr(Y_t = k) = P_{kt} = \frac{\exp(X_t\beta^k)}{\sum_{j=0}^{J} exp\ (X_t\beta^j)}.$$

Here, the parameter vector β^0=0. However, the vectors of others regressors are held constant and that additional parameter vector will be added for each additional alternative or categories. We calculate both relative risk ratios and marginal effects of variables under each marketing channel.

VI. RESULTS

Table 4 reports the results from the multinomial logit estimation and the relative risk ratios using the maximum likelihood estimation with a base category of no direct marketing. Thus, our analysis and interpretation of the results in Table 4 are relative to the basis which is no direct sale. As a result, the standard interpretation of the multinomial logit result is that for a unit change in the predictor variables, the logit of channel *m* relative to no direct sales is expected to change by its respective parameter estimate ceteris paribus.

Column (1, 2) in Table 4 is the estimation and risk ratios when farmers choose to directly sale to consumers from on-farm stores (Channel 1). Column (3, 4) in Table 4 is the estimation and risk ratios when farmers choose to sell to retail outlets (Channel 2) directly, and finally, column (5, 6) is the estimation and risk ratios when farmers choose to directly sell to consumers via multiple channels (Channel 3).

Regarding the choice of direct marketing channel relative to the base, gender of the operator has no influence on the channel choice. However, farm size has a significant impact on the choice of Channel 1 and Channel 3. If farm size were to increase by one acre, the multinomial log-odds for the Channel (1) relative to the base would be expected to decrease total sale from direct marketing of sale by 0.014 unit while holding all other variables in the model constant. Similarly, if farm size were to increase by one acre, the multinomial log-odds for the Channel (3) relative to the base would be expected to decrease total sale from direct marketing by 0.06 unit while holding all other variables in the model constant. Farmers located in the Northeast are more likely to engage in direct marketing than farmers who are located in the Midwest, Pacific, and Mountain states in the United States. Boys and Blank (2016) argue that these findings are partially due to variations in agricultural production between regions in the United States, which also highlights the global impact of consumer's heterogeneity in food choices. They also indicate that majority of the consumers who engage in the purchase of local food have at least a college degree or female consumers, or household with children. What is also significant in our findings this regional dummy is significant for all three channels. In the analysis, we looked at how farmer's financial variables affect the choice of direct marketing via the different channels and found that farmers' gross income does affect the choice of Channel 1. If farmers' gross income from the direct sales were to increase by one unit, the multinomial log-odds for the Channel (1) relative to the base would be expected to increase while holding all other variables in the model constant. However, farmers

who are located in the Mountain states prefer to sell via multiple channels like Channel 3. For farmers located in the Mountain states, the multinomial log-odds for the Channel (3) relative to the base would be expected to increase total sale from direct marketing by 1.006 units while holding all other variables in the model constant.

We perform the Wald test to see if outcome channels can be collapsed especially for Channel 3 where the null hypothesis is Ho: All coefficients except intercepts associated with channel 3 are 0 (i.e., categories can be collapsed). We found that when we join the results across channels LR test and Wald test produce similar results. For all combinations of categories especially in multiple channels where farmers combine both retail outlets and farmers markets and farm store, we reject the hypotheses that our variables do not differentiate between categories. So, our combination of both Channels 1 and 2 is appropriate and make a significant difference. We also test that the coefficients of the variables included in the estimations presented in Table 6 are zeros or can be dropped from the regression. We perform both the likelihood ratios tests and the Wald tests, and we found that the results are the same. We found that variables coefficients are significant across all outcome categories expect when farmers are located in the Pacific region.

We also test for independent of channel choices or categories using the Hausman test which is computed using no direct sale channel or base channel which is the most common choice observed in the data. We fail to reject the null hypothesis that the odds (Channel 1 versus Channel 2) are independent of Channel 3 at a 5 percent level

of significance. The value of the Hausman and McFadden (HM) tests are not significant indicating that the IIA assumption has not been violated and that difference in coefficients is systematic, consistent, and asymptotically distributed chi-square with 9 degrees of freedom under the null with 9 degrees of freedom. We also notice that the value of the HM test for total employment, total annual payroll, and total number of establishment relative to direct sale are not negative except for the Midwest and Mountain which throw a nuance in the validity of the test, but we don't the impact if we were the increase the sample size.

Table 4 presents the marginal effects from the multinomial regression model. We find that holding everything else constant at their means, a one unit increase in farm size decreases the likelihood of choosing Channel 1 by a minuscule fraction. Similarly, if a farm is located in one of the Northern states, the likelihood of choosing Channel 1 increases by 0.0367 units. However, the likelihood to engage in direct sale using Channel 2 by a farmer in one of the Northern states increases by 0.0094 units.

For Channel 3, a one unit increase in farm size decreases the likelihood to choose this channel by a very small fraction holding everything else constant at their means. Moreover, the likelihood of engaging in direct sales using Channel 3 for farmers in one of the states in Northern states increases by 0.0108 units. This value is higher than for Channel 1 but lower than for the Channel 2. Moreover, regional dummy for Channel 3 is also significant for Pacific and the Mountain regions. Holding everything else constant at their means, the likelihood of choosing the Channel 3 by farmers in the Pacific states increases by 0.0058 units. Holding everything else constant at their

means farmers located in one of the Mountain states are likely to increase Channel 3 use by 0.0046 units. Recall, farmers located in the Southern region are the base for the regional dummies.

Table 5 evaluates the effect of covariates and presents the margins or relative probabilities of channel choice using only the estimated sample. We find that given the direct sales channels available to farmers in the United States, farmers are less likely to engage in direct marketing. The average probability of no direct sales is 96% ceteris paribus. The average probability of disengagement from direct marketing is on average is 94% higher than direct sales into Channel 1 which is direct sales using farmers markets and on-farm store. However, the average probability of choosing Channel 1 and Channel 2 are on average the same about 0.4%.

VII. CONCLUSIONS

We analyzed factors affecting different direct marketing channel choice for both crop and livestock products in the United States using a multinomial logit regression model. We found that while there is increasing interest in local foods and direct marketing, the share of sales through these channels is very small. Our analysis also indicates that marketing of food system is changing in the United States. Farmers located in the Northeast are more likely to engage in direct marketing than farmers located in other regions. Those choosing to sell in multiple channels are higher than those choosing one channel outlet in the Northern part of the United States. However,

the average probability of no direct sale in the United States is 93% higher than Channels 1-3 combined ceteris paribus.

We also find that small size farmers are more likely to get engaged in direct marketing of crop and livestock products. Small producers may have difficulty finding large buyers because of volume and quality of outputs they produce.

References

Angelo, B. E., Jablonski, B. B., and Thilmany, D. (2016). Meta-analysis of US intermediated food markets: measuring what matters. *British Food Journal*, *118*(5), 1146-1162.

Boys, K. A. and Blank, S. C. (2017). The Evolution of Local Foods: A Retrospective and Prospective Consideration. In J. Stanton and M. Lang (Eds.), The Meaning of Local Foods: A Food Marketing Perspective (forthcoming).

Brown, A. (2002). Farmers' market research 1940–2000: An inventory and review. *American Journal of Alternative Agriculture*, *17*(04), 167-176.

Brown, C., Gandee, J. E., and D'Souza, G. (2006). West Virginia farm direct marketing: A county level analysis. *Journal of Agricultural and Applied Economics*, *38*(03), 575-584.

Buhr, B. L. (2004). Case studies of direct marketing value-added pork products in a commodity market. *Review of Agricultural Economics*, *26*(2), 266-279.

Capstick, D. F. (1982). A study of direct marketing of farm produce in Arkansas. *Bulletin-Arkansas, Agricultural Experiment Station (USA)*.

Darby, K., Batte, M. T., Ernst, S., and Roe, B. (2008). Decomposing local: a conjoint analysis of locally produced foods. *American Journal of Agricultural Economics*, *90*(2), 476-486.

Detre, J. D., Mark, T. B., Mishra, A. K., and Adhikari, A. (2011). Linkage between direct marketing and farm income: a double-hurdle approach. *Agribusiness*, *27*(1), 19-33.

Gallons, J., Toensmeyer, U. C., Bacon, J. R., and German, C. L. (1997). An analysis of consumer characteristics concerning direct marketing of fresh produce in Delaware: a case study. *Journal of Food Distribution Research*, *28*(1), 98-106.

Gormley, C., and B. Murphy. (2010). Clustering Ranked Preference Data Using Sociodemographic Covariates. Editors S. Hess, and A. Daly, 1st Edition, pp. 543–569.

Gormley, I.C., and T.B. Murphy. (2008). A Mixture of Experts Model for Rank Data with Applications in Election Studies. *The Annals of Applied Statistics* 2(4):1452-1477.

Gourieroux, C., and Monfort, A. (1996). *Simulation-based econometric methods*. Oxford University Press.

Govindasamy, R., Zurbriggen, M., Italia, J., Adelaja, A. O., Nitzsche, P., and VanVranken, R. (1998). Farmers Markets: Consumer Trends, Preferences, and Characteristics. http://ageconsearch.umn.edu/bitstream/36722/2/pa980798.pdf. Accessed on March 20, 2017.

Guthman, J., Morris, A. W., and Allen, P. (2006). Squaring farm security and food security in two types of alternative food institutions. *Rural Sociology*, *71*(4), 662.

Hay, A. M., and Smith, R. H. (1980). Consumer welfare in periodic market systems. *Transactions of the Institute of British Geographers*, 29-44.

Hsiao, C., Q. Li, and J.S. Racine. (2007). A Consistent Model Specification Test with Mixed Discrete and Continuous Data. *Journal of Econometrics* 140(2):802-826.

Jacobs, R.A., M.I. Jordan, S.J. Nowlan, and G.E. Hinton. (1991). Adaptive Mixtures of Local Experts. *Neural Computation* 3(1):79-87.

Jarosz, L. (2008). The city in the country: Growing alternative food networks in Metropolitan areas. *Journal of Rural Studies*, *24*(3), 231-244.

Jordan, M.I., and R.A. Jacobs. (1994). Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation* 6(2):181-214.

Kohls, R. L., and Uhl, J. N. (2002). *Marketing of agricultural products* (No. Ed. 9). Prentice-Hall Inc.

Low, S. A., and Vogel, S. J. (2011). Direct and intermediated marketing of local foods in the United States. USDA Economics Research Service, Economic Research Report Number 118. https://www.ers.usda.gov/webdocs/publications/err128/8276_err128_2_.pdf. Accessed on March 1, 2017.

McLachlan, G.J. and D. Peel. 2000. Finite Mixture Models. Wiley, New York. MR1789474.

Metcalf, J. 1999. Dane County farmers' market offers diversity, quality. The Vegetable Growers News, Sparta Michigan. January, 62-63.

Monson, J., Mainville, D., and Kuminoff, N. (2008). The decision to direct market: an analysis of small fruit and specialty-product markets in Virginia. *Journal of Food Distribution Research*, *39*(2), 1-11.

Murteira, J., E. Ramalho, and J. Ramalho. (2013). Regression Analysis of Multivariate Fractional Data. Forthcoming *Econometric Reviews* DOI: 10.1080/07474938.2013.806849

Papke, L.E., and J.M. Wooldridge. (1996). Econometric Methods for Fractional Response Variables with an Application to 401(K) Plan Participation Rates." *Journal of Applied Econometrics Journal* 11(6):619–632.

Papke, L.E., and J.M. Wooldridge. (2008). Panel Data Methods for Fractional Response Variables with an Application to Test Pass Rates." *Journal of Econometrics* 145(1):121-133.

Park, T. A. (2009). Assessing the returns from organic marketing channels. *Journal of Agricultural and Resource Economics*, *34*(3), 483-497.

Park, T. (2015). Direct marketing and the structure of farm sales: An unconditional quantile regression approach. *Journal of Agricultural and Resource Economics*, *40*(2), 266-284.

Park, T., Mishra, A. K., & Wozniak, S. J. (2014). Do farm operators benefit from direct to consumer marketing strategies? *Agricultural Economics*, *45*(2), 213-224.

Ramalho, E.A., J.J.S. Ramalho, and J.M.R. Murteira. (2011). Alternative Estimating and Testing Empirical Strategies for Fractional Regression Models. *Journal of Economic Surveys* 25(1):19-68.

Sage, J. L. (2012). *A Geographic Exploration of the Social and Economic Sustainability of Farmers' Markets and the Rural Communities That Make Them Work* (Doctoral Dissertation, Washington State University).

Schneider, M. L., and Francis, C. A. (2005). Marketing locally produced foods: Consumer and farmer opinions in Washington County, Nebraska. *Renewable Agriculture and Food Systems*, *20*(04), 252-260.

Teller, C., and Reutterer, T. (2008). The evolving concept of retail attractiveness: what makes retail agglomerations attractive when customers shop at them? *Journal of Retailing and Consumer Services*, *15*(3), 127-143.

Uematsu, H., and Mishra, A. K. (2011). Use of direct marketing strategies by farmers and their impact on farm business income. *Agricultural and Resource Economics Review*, *40*(1), 1-19.

Wolf, M. M. (1997). A target consumer profile and positioning for promotion of the direct marketing of fresh produce: a case study. *Journal of Food Distribution Research*, *28*(3), 11-17.

Table 1: Variable Descriptions

Variables	Descriptions Marketing
Dependent Variables	
NODIRECT	Percent of sale to No-direct marketing
DIRECT-FARMER	Percent of sale directly to consumers at farmers market or on farm store, road side stand
DIRECT-RETAIL	Percent of sale directly to local retail outlet such as restaurant or grocery store that sold directly to individual consumers
DIRECT-OTHER	Direct to regional distributor or local institution
Independent Variables	
OP_GEN	Gender of operator (= 1 if male, 0 otherwise)
GFI	Gross farm income
OPMH	Operator working distance miles from home
YOB	Year operator began
TACRES	Total acres farmed
WAGL	Wage expense for others labor
PACIFIC	Farm is located in Pacific states (1 = yes, 0 otherwise)
MOUNTAIN	Farm is located in Mountain states (1 = yes, 0 otherwise)
MIDWEST	Farm is located in Midwest states (1 = yes, 0 otherwise)
NORTHEAST	Farm is located in Northeastern states (1 = yes, 0 otherwise)
TOTANPAY	Total annual payroll for all four NAICS industry within the county
TOTEMP3	Total number of employment in four NAICS industries within the county
TOTESTAB	Total number of establishments of all four industries within the county

Note: Data are from the USDA's Agricultural Resource Management Survey and the United States Census Bureau. The NAICS industries included in the study are 4244 (Grocery and related product merchant wholesalers), 4451 (Grocery stores), 7225 (Restaurant and other eating places), and 4245 (Farm product raw material merchant wholesalers).

Table 2: Descriptive Statistics

Table 3: Coefficients of Factors Affecting Marketing Channel Choice Obtained from a Multinomial Logit Regression Model

Variables	Channel 1		Channel 2		Channel 3	
	Coefficients	RRR	Coefficients	RRR	Coefficients	RRR
GFI	1.62e-08**	1.00	-3.82e-09	1.00	-4.39e-08	1.00
	(8.28e-09)	(0.00)	(4.40e-08)	(0.00)	(1.21e-07)	(0.00)
OP_GEN	0.0954	1.10**	-0.111	0.89	-0.0277	0.97
	(0.254)	(0.28)	(0.604)	(0.54)	(0.531)	(0.52)
WAGL	8.60e-08	1.00	8.12e-09	1.00	1.34e-07	1.00
	(1.45e-07)	(0.00)	(4.19e-07)	(0.00)	(6.94e-07)	(0.00)
TACRES	-0.000148**	1.00	4.61e-06	1.00	-0.000601**	1.00
	(5.82e-05)	(0.00)	(1.03e-05)	(0.00)	(0.000297)	(0.00)
YOB	8.90e-05	1.00	0.000367*	1.00	0.000181	1.00
	(0.000138)	(0.00)	(0.000217)	(0.00)	(0.000312)	(0.00)
OPMH	0.00232	1.00	0.00193	1.00	-0.00473	1.00
	(0.00182)	(0.00)	(0.00409)	(0.00)	(0.0115)	(0.01)
NORTHEAST	1.696***	5.45***	2.106***	8.22***	2.448***	11.55***
	(0.171)	(0.93)	(0.368)	(3.02)	(0.368)	(4.25)
PACIFIC	0.110	1.12***	0.631	1.87	1.265***	3.54***
	(0.238)	(0.27)	(0.450)	(0.85)	(0.451)	(1.59)
MIDWEST	-0.420	0.66	-0.361	0.70	-0.680	0.51
	(0.319)	(0.21)	(0.746)	(0.52)	(1.034)	(0.52)
MOUNTAIN	-0.0305	0.97	0.671	1.95	1.006*	2.73
	(0.303)	(0.29)	(0.526)	(1.02)	(0.565)	(1.54)
TOTEMP	6.71e-05	1.00	2.55e-05	1.00	-5.87e-05	1.00
	(6.00e-05)	(0.00)	(8.34e-05)	(0.00)	(0.000159)	(0.00)
TOTANPAY	-3.07e-06	1.00	1.74e-08	1.00	3.38e-06	1.00
	(2.46e-06)	(0.00)	(3.60e-06)	(0.00)	(5.96e-06)	(0.00)
TOTESTAB	0.000543	1.00	8.59e-05	1.00	-0.00174	1.00
	(0.000974)	(0.00)	(0.00188)	(0.00)	(0.00407)	(0.00)

CONSTANT	-3.993***	0.02***	-5.796***	0.00	-5.550***	0.00
	(0.290)	(0.01)	(0.678)	(0.00)	(0.640)	(0.00)
OBS.	9,992	9,992	9,992	9,992	9,992	9,992

Note: This is a Direct Marketing sale of both grains and livestock by farmers. Column (1) is Channel 1; Column (2) is Channel 2, and column (3) is Channel 3. Channel 0 of no direct sale is the base group. Robust Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. RRR are the relative risk ratio relative to its channel.

Table 4: Marginal Effects

Table 5: Predicted probabilities of Channel Choice

Channels	Margin	Delta-method Std. Err.	Z	P>z
Channel 1	0.0232186	0.001492	15.56	0
Channel 2	0.0048038	0.0006878	6.98	0
Channel 3	0.0047038	0.0006813	6.9	0
No Direct marketing	0.9672738	0.0017518	552.17	0